**T.C.**

**BİLECİK ŞEYH EDEBALİ ÜNİVERSİTESİ**

**İKTİSADİ VE İDARİ BİLİMLER FAKÜLTESİ**

**YÖNETİM BİLİŞİM SİSTEMLERİ**

****

**Yudum AKKAYA**

**69985040494**

**VERİ MADENCİLİĞİ**

**Dr. Öğr. Üyesi Nur Kuban TORUN**

**Tüketicinin Belli Maaş Düzeyinde Bir Ürünü Satın Alıp Almayacağının Tahmininde Sınıflandırma Algoritmalarının Karşılaştırılması**

1. **Giriş**

Tüketici genel olarak ekonomik değeri olan ve bir satıcı tarafından satılan mal ve hizmetleri satın alma isteği ve satın alabilecek ekonomik gücü olan kişidir. İstek ya da ekonomik güçten birisi olmadığı zamanda potansiyel tüketiciden söz edilebilir. Örneğin bir bebek otomobil satıcısı için, karnı tok olan birisi lokanta için, potansiyel tüketici olabilir. Genel olarak ticaret ile uğraşanlar tüketici kelimesinin yerine müşteri kelimesini daha çok “satın alan kişi” anlamında kullanmaktadır. Pazarlama literatürüne sonradan girmiş ve hem tüketici hem de müşteri kelimesinin yerine kullanılan bir kelime de alıcıdır. Pazarlama, tüketici davranışının satın alınmadan önceki süreçten başlayarak, satın almadan ve hatta tüketimden sonraki süreçte olanları da ele alan bir yaklaşıma sahiptir.

Tüketici davranışının iki temel boyutu bulunmaktadır. Bu boyutlar birbirini tamamlamakta ve bir bütün oluşturmaktadırlar. Birinci boyut tüketicinin karar alma sürecidir. İkinci önemli boyut ise satın alma davranışıdır. Satın alma kararının zihinde oluşması rasyonel olabildiği gibi duygusal yanları da bulunabilmektedir.

Ekonomik Faktörler : Tüketicilerin gelirleri ile gelirlerini elde etme biçimleri tüketimlerine de yansımaktadır. Zaten tüketim ve tüketici davranışı gelirin doğal bir sonucudur. Çünkü geliri olmayan birisinin tüketici olması düşünülemez. Gelirin varlığı kadar büyüklüğü ve kaynağı da tüketicinin harcamaları bakımından önemlidir.

Bir ülkedeki bireylerin ortalama zenginlik göstergesi olarak kabul edilebilecek kişi başına gelir rakamları, tüketimi etkileyen önemli bir kriterdir. Bir mal ya da hizmet için bireysel tüketim amacıyla talep edilen miktarlar toplanarak ülke için talep tahmin edilebilir. Kullanılabilir kişisel gelir, tüketicilerin vergileri ödedikten sonra ellerinde kalan ve harcayabilecekleri gelirdir.

Kişi başına gelirle birlikte, bireylerin satın alma güçlerini oluşturan “kullanılabilir kişisel gelir” ile “isteğe bağlı satın alma gücü” rakamları da pazarlama yönetimi açısından izlenmesi gereken ekonomik faktörlerdendir. Kullanılabilir kişisel gelir kişisel gelirden kişilerin ödedikleri gelir vergisi çıktıktan sonra kalan gelirdir ve pazarlama yöneticileri açısından kişi başına milli gelirden daha önemli bir ölçüttür.

**1.1 Veri Madenciliği Nedir?**

Büyük veri tabanlarında ilginç ve değerli olan bilgiyi algılamak ve erişmek oldukça zordur. Veri Tabanında Bilgi Keşif sürecinin aşamaları (Knowledge Discovery in Databases) bu değerli, önceden bilinmeyen, kullanılabilir olan bilgiye belirli yöntemler uygulayarak ulaşmada çok büyük rol oynamaktadır. Veri madenciliğinin uygulanabilmesi için kullanılan algoritma bize veri tabanlarında bilgi keşfi sürecinin gerçekleşmesini sağlamaktadır. Bu süreç içerisinde modelin uygulanacağı verilerin özelliklerinin çok iyi bilinmesi gerekmektedir.

Veri madenciliği bir süreçtir. Veri yığınları arasında soyut kazılar yaparak veriyi ortaya çıkarmanın yanı sıra, bilgi keşfi sürecinde örüntüleri ayrıştırarak süzmek ve bir sonraki adıma hazır hale getirmek de bu sürecin bir parçasıdır.

Üzerinde inceleme yapılan işin ve verilerin özelliklerinin bilinmemesi durumunda ne kadar etkin olursa olsun hiç bir veri madenciliği algoritmasının fayda sağlaması mümkün değildir. Bu sebeple, veri madenciliği sürecine girilmeden önce, başarının ilk şartı, iş ve veri özelliklerinin detaylı analiz edilmesidir.

Veri madenciliği sürecinde izlenen adımlar genellikle şöyledir :

1. Problemin tanımlanması,
2. Verilerin hazırlanması,
3. Modelin kurulması ve değerlendirilmesi,
4. Modelin kullanılması,
5. Modelin izlenmesi.

**Problemin tanımlanması**

Bu aşama veri madenciliği sürecinin en önemli aşamasıdır. Araştırma probleminin (konusunun) tanımlanması aşaması araştırmanın amacını, mevcut durumun değerlendirilmesini, veri madenciliğinin amaçlarını ve proje planlama sürecinin belirlenmesini kapsamaktadır.

**Verilerin hazırlanması**

Modelin kurulması aşamasında ortaya çıkacak sorunlar, bu aşamaya sık sık geri dönülmesine ve verilerin yeniden düzenlenmesine sebep olacaktır.

Bu aşamada firmanın mevcut bilgi sistemleri üzerinde ürettiği sayısal bilginin iyi analiz edilmesi, veriler ile mevcut iş problemi arasında ilişki olması gerektiği unutulmamalıdır.

Proje kapsamında kullanılacak sayısal verilerin, hangi iş süreçleri ile yaratıldığı da bu veriler kullanılmadan analiz edilmelidir, bu sayede analist veri kalitesi hakkında fikir sahibi olabilir.

Veri kalitesi, veri madenciliğinde anahtar bir konudur. Veri madenciliğinde güvenilirliğin artırılması için, veri ön işleme yapılmalıdır. Aksi halde hatalı girdi verileri kullanıcıyı hatalı çıktıya götürecektir. Veri ön işleme, çoğu durumlarda yarı otomatik olan ve yukarıda da belirtildiği gibi zaman isteyen bir veri madenciliği aşamasıdır. Verilerin sayısındaki artış ve buna bağlı olarak çok büyük sayıda verilerin ön işlemeden geçirilmesinin gerekliliği, otomatik veri ön işleme için etkin teknikleri önemli hale getirmiştir.

Verilerin hazırlanması şu aşamalarından oluşmaktadır:

* Toplama,
* Değer biçme,
* Birleştirme ve temizleme,
* Örneklem seçimi,
* Dönüştürme.

**Modelin kurulması ve değerlendirilmesi**

Tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu sebeple veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar yinelenen bir süreçtir. Model kuruluş süreci, denetimli ve denetimsiz öğrenmenin kullanıldığı modellere göre farklılık göstermektedir.

Örnekten öğrenme olarak da isimlendirilen denetimli öğrenmede, bir denetçi tarafından ilgili sınıflar önceden belirlenen bir kritere göre ayrılarak, her sınıf için çeşitli örnekler verilir. Sistemin amacı verilen örneklerden hareket ederek her bir sınıfa ilişkin özelliklerin bulunması ve bu özelliklerin kural cümleleri ile ifade edilmesidir. Öğrenme süreci tamamlandığında, tanımlanan kural cümleleri verilen yeni örneklere uygulanır ve yeni örneklerin hangi sınıfa ait olduğu kurulan model tarafından belirlenir.

Denetimsiz öğrenmede, kümeleme analizinde olduğu gibi ilgili örneklerin gözlenmesi ve bu örneklerin özellikleri arasındaki benzerliklerden hareket ederek sınıfların tanımlanması amaçlanmaktadır.

Denetimli öğrenmede seçilen algoritmaya uygun olarak ilgili veriler hazırlandıktan sonra, ilk aşamada verinin bir kısmı modelin öğrenilmesi, diğer kısmı ise modelin geçerliliğinin test edilmesi için ayrılır. Modelin öğrenilmesi öğrenim kümesi kullanılarak gerçekleştirildikten sonra, test kümesi ile modelin doğruluk derecesi belirlenir.

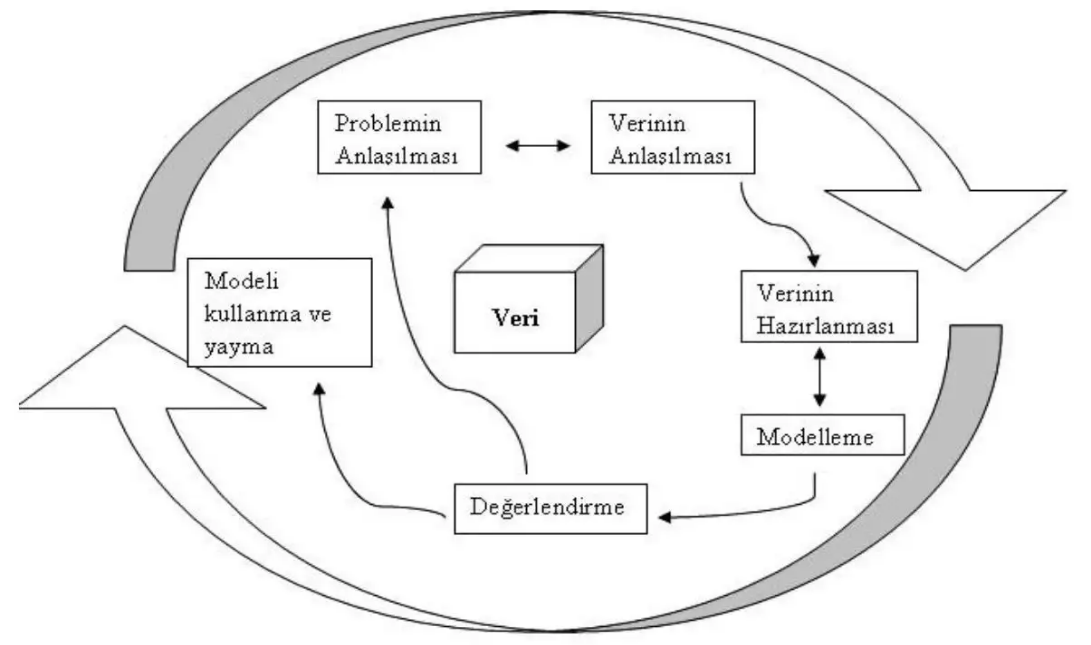
Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik testidir. Bu yöntemde tipik olarak verilerin %5 ile %33 arasındaki bir kısmı test verileri olarak ayrılır ve kalan kısım üzerinde modelin öğrenimi gerçekleştirildikten sonra, bu veriler üzerinde test işlemi yapılır. Bir sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile ise doğruluk oranı hesaplanır.

**Modelin kullanılması**

Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir. Kurulan modeller, kullanıldığı alanda amaçlanan problem çözümleri için anlamlı örüntüler ve değerlendirmeler oluşturmalıdır.

**Modelin izlenmesi**

Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve yeniden düzenlenmesini gerektirecektir. Tahmin edilen ve değişkenler arasındaki farklılığı gösteren grafikler model sonuçlarının izlenmesinde kullanılan yararlı bir yöntemdir.



Hazırladığım bu ödevde kullandığım veri seti tüketicilerin belli bir maaş düzeyinde herhangi bir ürünü satın alır mı? Bunun cevabını alabilmek. Veri madenciliği dersi kapsamında öğrendiğimiz Tahmin Edici Yöntemlerin Sınıflandırma başlığı altında bulunan Karar Ağaçları ve Naïve Bayes yöntemleriyle veri setimizin doğruluk oranlarını, R program yardımıyla bulup sonrasında bu iki yöntemin doğruluk oranlarını karşılaştırıp veri setimiz için hangi yöntem daha doğru onu bulacağız.

1. **Verinin Elde Edilmesi**

Veri herhangi bir tüketici kitlesinin maaş faktörüne göre bir ürünü satın alıp almayacağıyla ilgilidir. Veri setimi uzun araştırmalarım sonucu GitHub adlı platformdan buldum. Veri setim 5 öznitelik olmak üzere toplam 400 veriden oluşmaktadır.

1. **Metodoloj**
   1. **Karar Ağaçları Sınıflandırma**

Karar ağaçları, sınıflandırma ve tahmin için sıkça kullanılan bir veri madenciliği yaklaşımıdır. Sinir ağları gibi diğer metodolojilerin de sınıflandırma için kullanılabilmesine rağmen, karar ağaçları, kolay yorumu ve anlaşılabilirliği açısından karar vericiler için avantaj sağlamaktadır .

Karar ağaçları;

• Düşük maliyetli olması,

• Anlaşılmasının, yorumlanmasının ve veri tabanları ile entegrasyonun kolaylığı,

• Güvenilirliklerinin iyi olması

gibi nedenlerden ötürü en yaygın kullanılan sınıflandırma tekniklerinden biridir. Karar ağacı tekniğini kullanarak verinin sınıflanması, öğrenme ve sınıflama olmak üzere iki basamaklı bir işlemdir. Öğrenme basamağında önceden bilinen bir eğitim verisi, model oluşturmak amacıyla sınıflama algoritması tarafından analiz edilir. Öğrenilen model, sınıflama kuralları veya karar ağacı olarak gösterilir. Sınıflama basamağında ise test verisi, sınıflama kurallarının veya karar ağacının doğruluğunu belirlemek amacıyla kullanılır. Eğer doğruluk kabul edilebilir oranda ise kurallar, yeni verilerin sınıflanması amacıyla kullanılır.

Eğitim verisindeki hangi alanların hangi sırada kullanılarak ağacın oluşturulacağı belirlenmelidir. Bu amaçla en yaygın olarak kullanılan ölçüm, Entropi ölçümüdür.

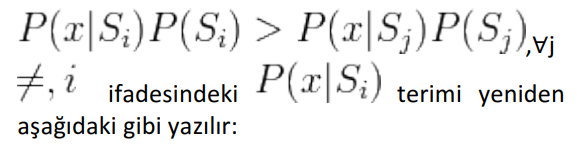
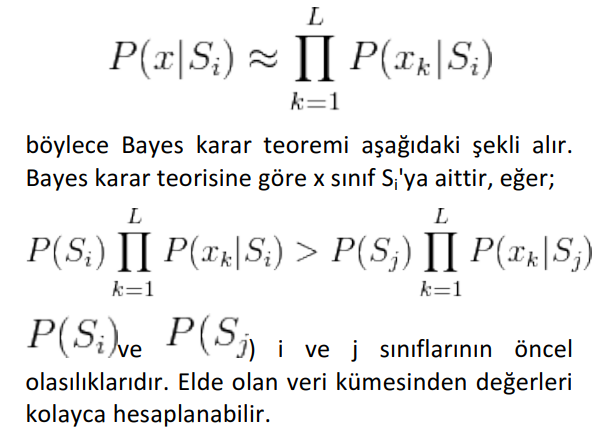
Entropi ölçüsü ne kadar fazla ise o alan kullanılarak ortaya konulan sonuçlar da o oranda belirsiz ve kararsızdır. Bu nedenle, karar ağacının kökünde Entropi ölçüsü en az olan alanlar kullanılır. Karar ağaçlarında kullanılan birçok algoritma mevcuttur. ID3, C4.5, C5.0, CART, CHAID ve QUEST bunlara örnek olarak gösterilebilir.

**C4.5 ve C5.0 Algoritmaları:** En yaygın kullanılan karar ağacı algoritması Quinlan’ın ID3 algoritmasının geliştirilmiş hali olan C4.5 algoritmasıdır. C5.0 algoritması ise C4.5’in geliştirilmiş hali olup, özellikle büyük veri setleri için kullanılmaktadır. C5.0 algoritması, doğruluğu arttırmak için boosting algoritmasını kullandığından, boosting ağaçları olarak da bilinir. C5.0 algoritması C4.5’e göre çok daha hızlı olup, hafızayı daha verimli kullanmaktadır. Her iki algoritmanın sonuçları aynı olsa da C5.0 biçim olarak daha düzgün karar ağaçları elde etmemizi sağlamaktadır.

* 1. **Naïve Bayes Sınıflandırma**

Naïve Bayes (TAN) sınıflandırma en etkili ve etkin bir sınıflandırma algoritmasıdır. Bayesyan ağların özel bir vakasını içermektedir. Sınırlandırılmamış bayesyan ağların yapısı ve parametreleri iyileştirmeler için anlamlı sonuçlar vermektedir. Tan, Friedman (1997) tarafından kıyaslama yapmaya yarayan büyük verilerden hızlı sonuç almak amacıyla geliştirilmiştir. Bayesyan sınıflandırma bir gruba veya sınıfa ait belirli bir parçayı tahmin etmekte kullanılmaktadır. Bu teknik, hızlı olmasından dolayı ve büyük veri setlerinde tutarlı sonuçlar alınmasından dolayı tercih edilmektedir. Dataları tahmin etme sırasında TAN, oldukça küçük bir eğitim veri setine ihtiyaç duymakta ve gerçek ve soyut verilerin sınıflandırılmasında başarı sağlamaktadır. Bayesyan sınıflandırmanın en büyük dezavantajı, bazı durumlarda ele aldığı sorunların gerçek problemler olmamasıdır. Bu teknik, sadece istatistik alanına değil; her hangi seçilen bir alanda rahatlıkla uygulanabilmektedir.

Verilen bir x'in ( x = [x(1), x(2), . . . , x(L)]T ∈ RL ) sınıf Si'ye ait olup olmadığına karar vermek için kullanılan yukarıda formüle edilen Bayes karar teoreminde istatistik olarak bağımsızlık önermesinden yararlanılırsa bu tip sınıflandırmaya Naive Bayes sınıflandırılması denir. Matematiksel bir ifadeyle:

****

Naive Bayes sınıflandırıcının kullanım alanı her ne kadar kısıtlı gözükse de yüksek boyutlu uzayda ve yeterli sayıda veriyle x'in (nicelik kümesi) bileşenlerinin istatistik olarak bağımsız olması koşulu esnetilerek başarılı sonuçlar elde edilebilinir.

1. **Kullanılan Modelin Ölçütleri**

Modelin başarısını ölçmek için kullanılan temel kavramlar hata oranı, kesinlik, duyarlılık ve Fölçütüdür. Modelin başarısı, doğru sınıfa atanan örnek sayısı ve yanlış sınıfa atılan örnek sayısı nicelikleriyle alakalıdır (Coşkun ve Baykal, 2011).

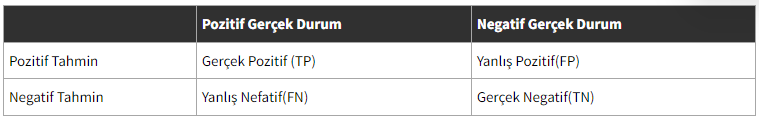
*Doğruluk oranı:* Doğru sınıflandırılmış örnek sayısının toplam örnek sayısına oranıdır.

*Hata oranı:* Yanlış sınıflandırılmış örnek sayısının toplam örnek sayısına oranıdır.

*Kesinlik oranı:* Pozitif olarak tahmin edilen doğru örnek sayısının, pozitif olarak tahminlenen tüm örnek sayısına oranıdır.

*Duyarlılık oranı:* Doğru sınıflandırılmış pozitif örnek sayısının; toplam pozitif örnek sayısına oranıdır.

*Özgüllük(belirlilik oranı):* doğru negatiflerin toplam yanlışlara oranıdır.

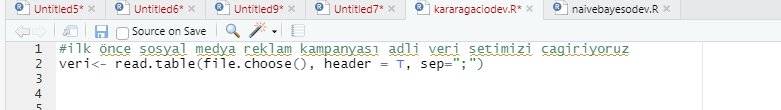


1. **Uygulama**

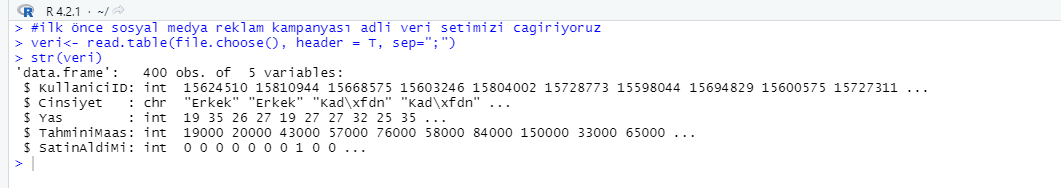
Bu çalışmada, tüketicilerin maaş faktörünü göz önüne alarak bir ürünü alıp almayacağının kararını tahminlemek için sınıflandırma algoritmalarından karar ağaçları ve naive bayes yöntemleri kullanılmıştır. Yöntemler R programında çözülmüştür. Veri seti 400 gözlem değeri ve 5 adet değişkenden oluşmaktadır. Veriler analiz için uygun hale getirilmiştir. Bunun için veri türü chr olan veriler factore çevrilmiştir. 1-0 içeren binary kategorik verilerde factore dönüştürülmüş ve iki analiz de bu şekilde çözümlenmiştir.

* 1. **Karar Ağaçları Sınıflaması**

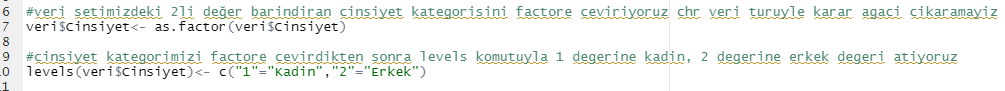
İlk önce csv formatında exele aktardığımız ve düzenlediğimiz veri setimizi R’a çağırıyoruz.

****

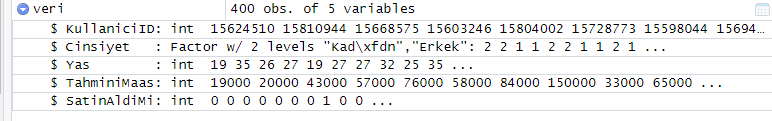
#k



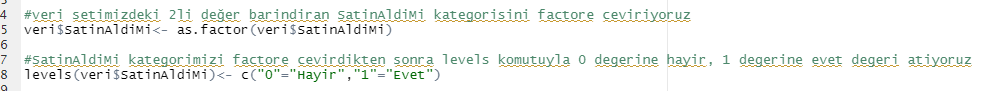
Cinsiyet faktörü gördüğünüz gibi chr olarak gelmiş biz karar ağacında chr faktörüyle analiz yapamıyoruz bu yüzden faktöre dönüştüreceğiz ve sonrasında da 1’e Kadın, 2’ye Erkek değeri atayacağız.

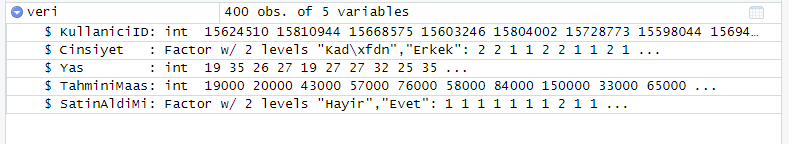


Çıktısı:



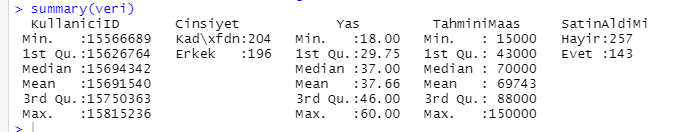
Satın Alma faktörü integer olarak gelmiş ama biz çıktıları daha net analiz edebilelim diye faktöre dönüştüreceğiz ve 0’a Hayır, 1’e Evet değeri atayacağız.





Verilerimizin istatiksel sonuçlarını görmek istersek summary komutunu kullanabiliriz.

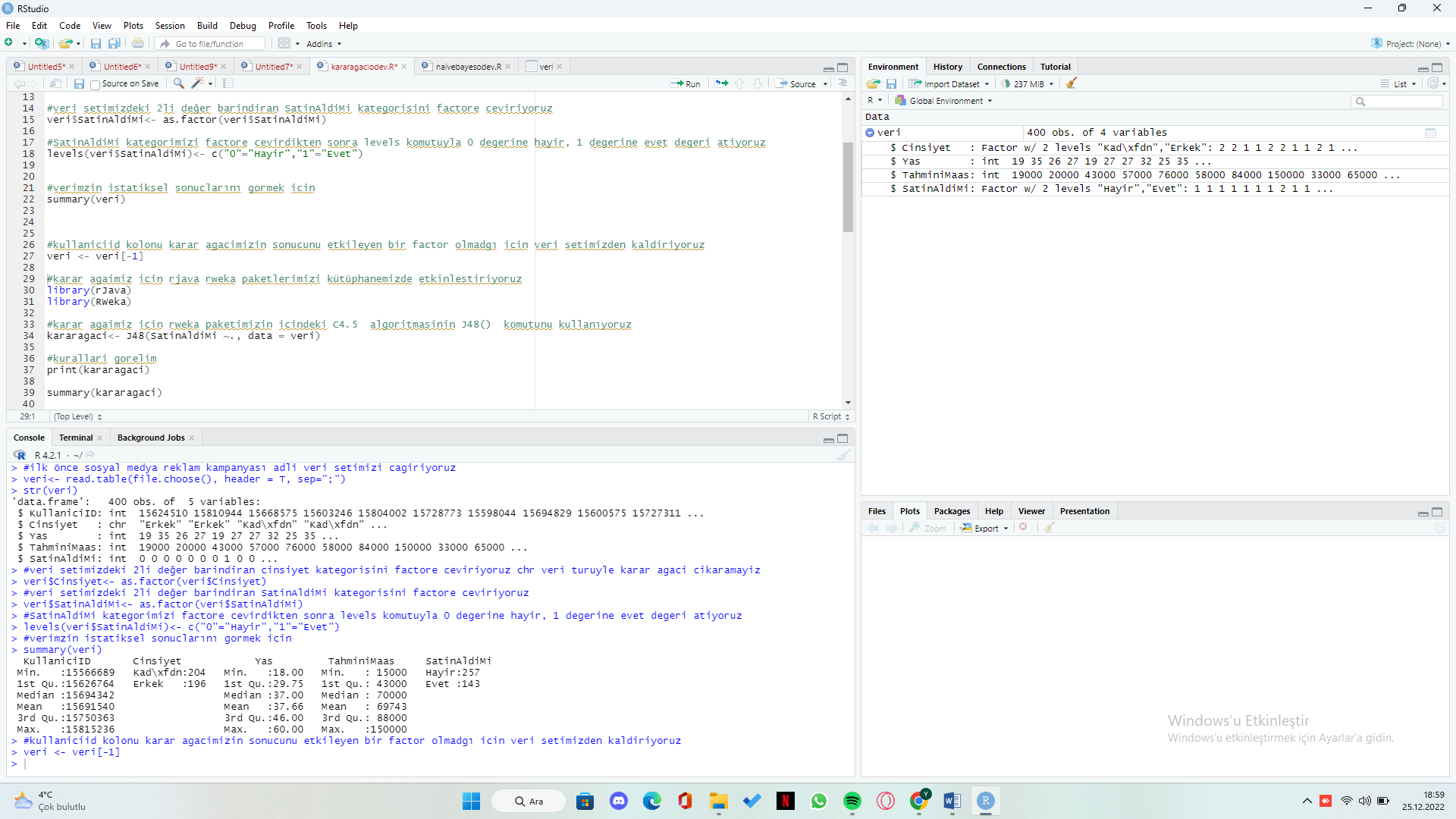




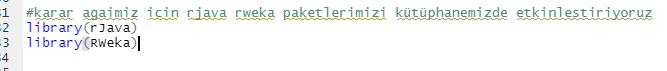
İlk sütun Kullanıcı Id sütunu olduğu için ve satın alma karar ağacını çıkarırken etkileyen bir factor olmadığı için siliyoruz.



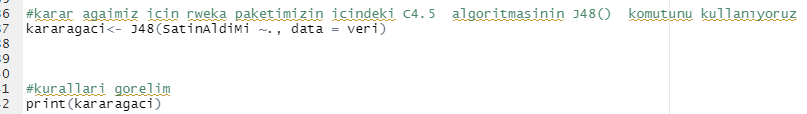
Silindiğini şu şekilde gösterelim:



R programında kütüphanemize paket yüklemek için ınstall.packages() komutunu kullanırız. Karar Ağacı yöntemini kullanırken bize gerekli olan paketler “rJava” ve “rWeka” paketleridir. Ben daha önce kütüphaneme eklediğim için sadece library() komutuyla çağırıyorum.



Karar ağacı için rWeka paketinin içerisindeki C4.5 algoritmasının J48() komutunu kullanıyoruz. Sonrasında da kurallarımızı görmek için print() fonksiyonunu kullanıyoruz.

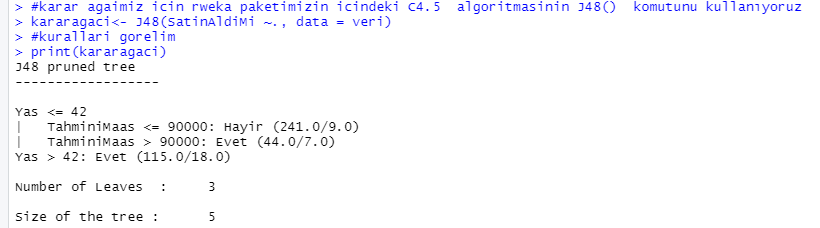


Çıktımızda kurallarımız görülmektedir. Kurallar:

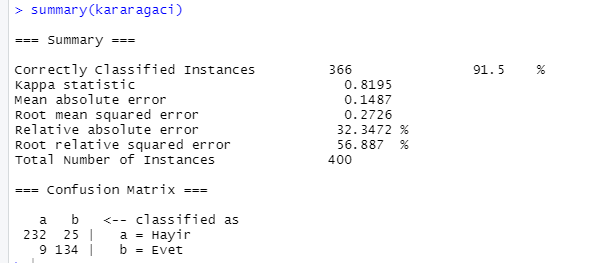
Eğer yaş 42’ye eşit veya küçükse ve tahmini maaşı 90000’e eşit veya küçükse HAYIR almaz.

Eğer yaş 42’ye eşit veya küçükse ve tahmini maaşı 90000’den büyükse EVET alır.

Eğer yaş 42’den büyükse tahmini maaşa bakılmaksızın EVET alır.

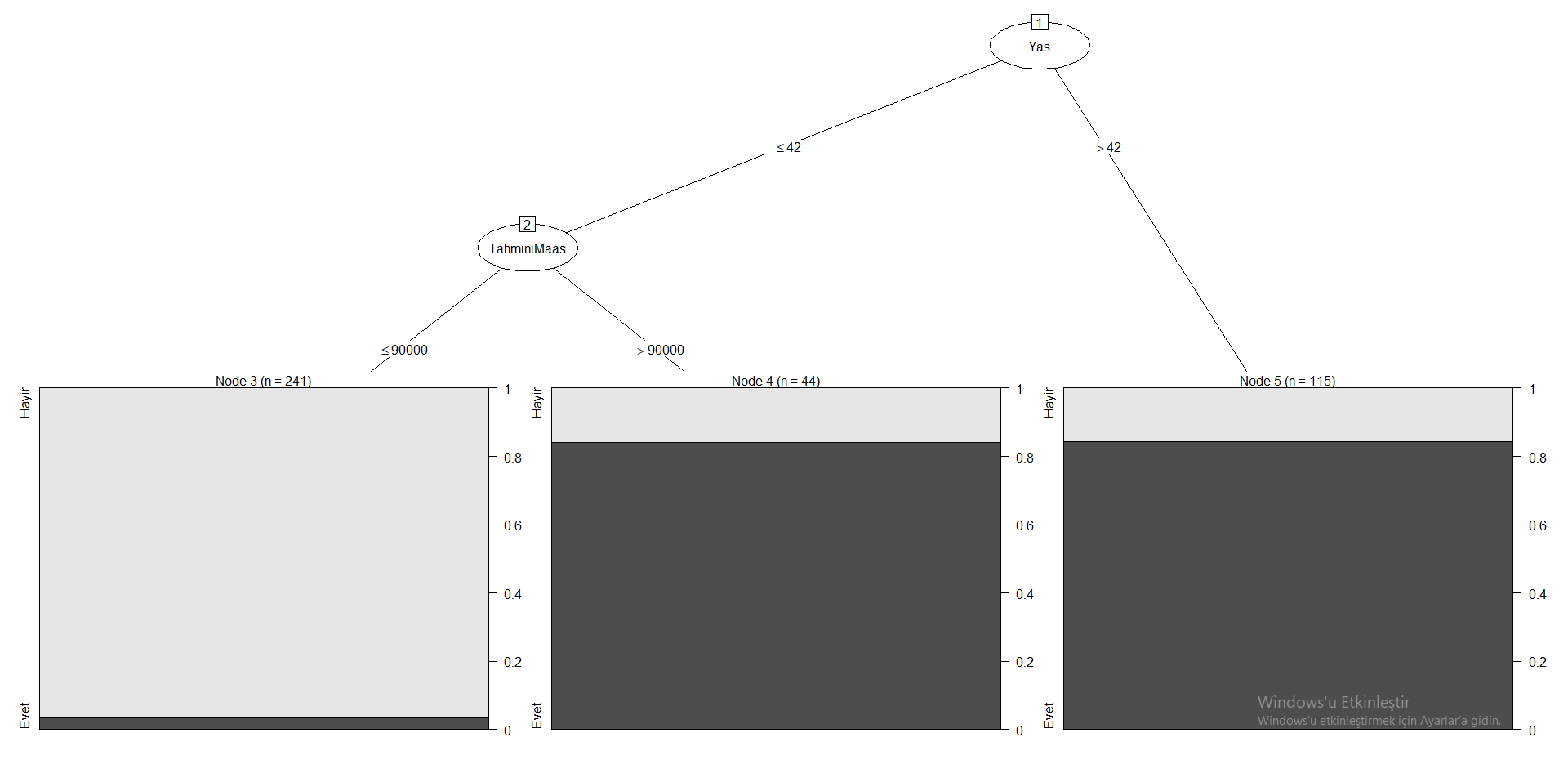


Summary() komutuyla çağırdığımızda ise veri setimizin %91.5 oranında doğru olduğunu gösteriyor.



Plot fonksiyonuyla ağacımızın grafiğini çizelim.

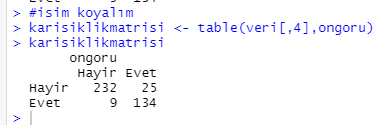




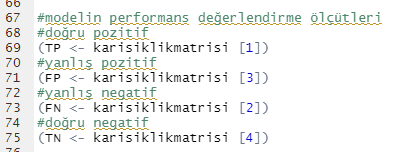
Modelimizin karışıklık matrisi görelim.



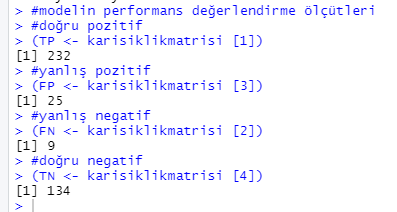
Çıktısı:



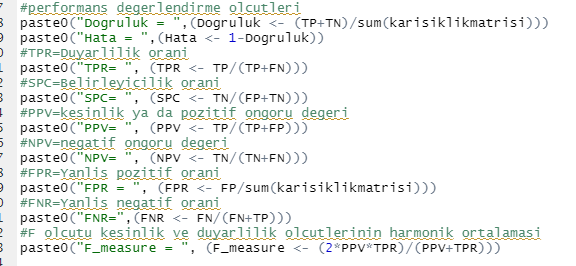
Modelimizin performans değerlendirme ölçütleri:



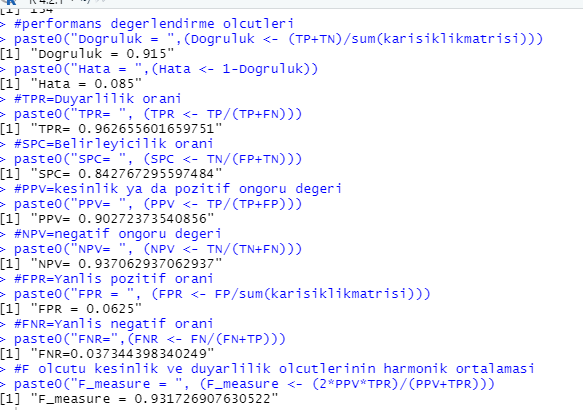
Çıktıları:



Modelimizin performans değerlendirme ölçütleri:



Çıktısı:

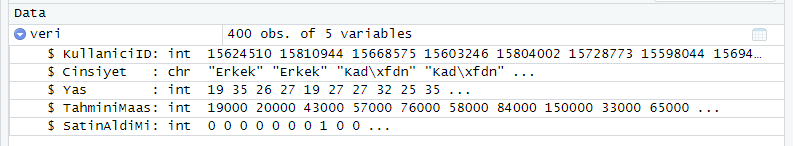


* 1. **Naïve Bayes Sınıflaması**

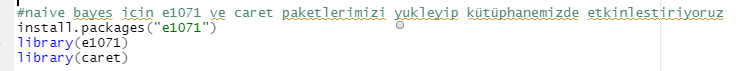
İlk önce csv formatında exele aktardığımız ve düzenlediğimiz veri setimizi R’a çağırıyoruz.

****

Verilerimiz:



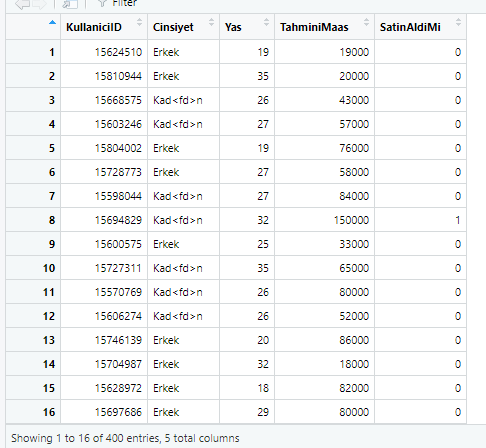
Daha sonra Naive Bayes için gerekli olan “Caret” ve “e1071” adlı paketlerimizi kütüphanemize yükleyip library() komutuyla çağırıyoruz.



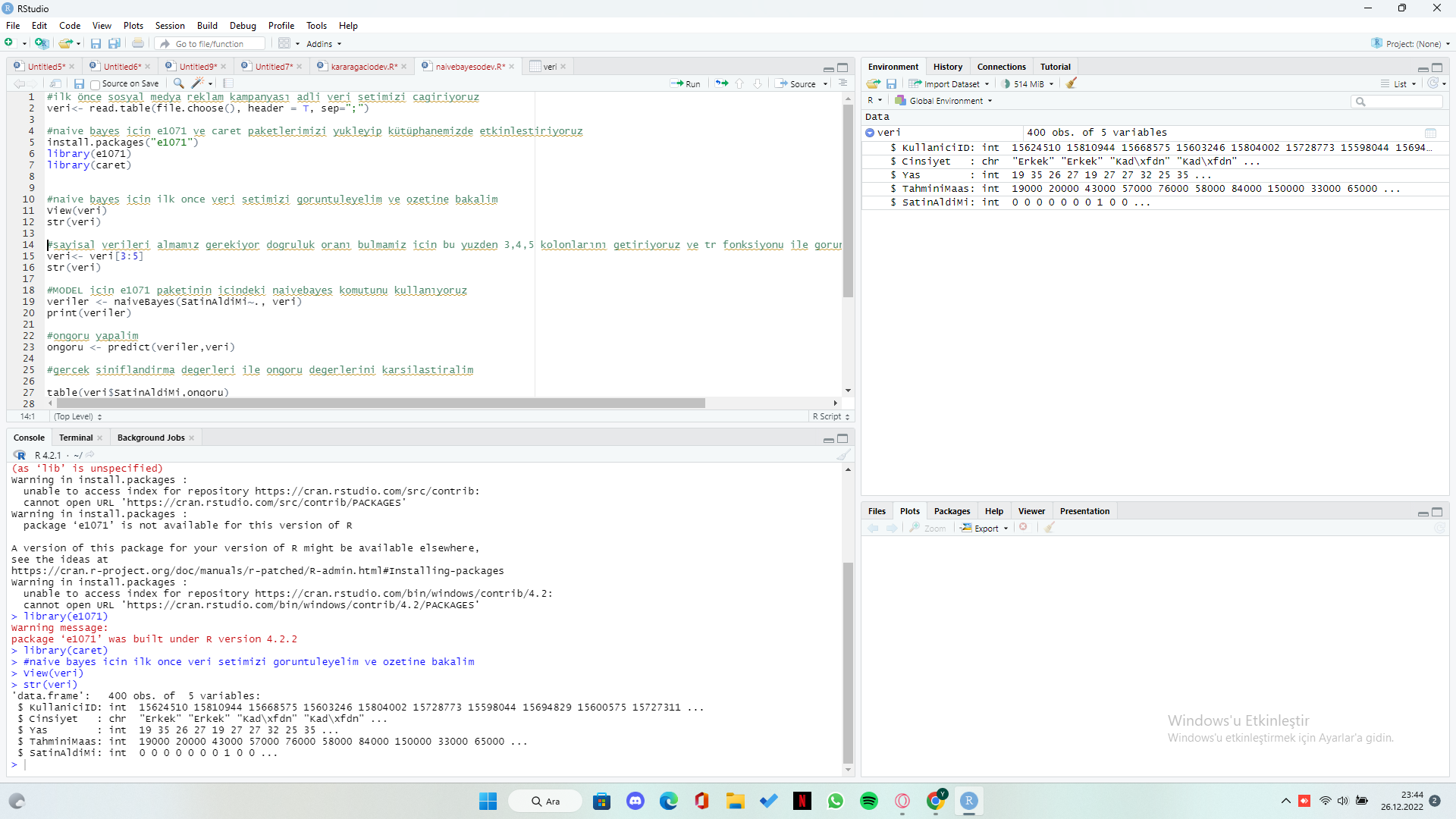
Öncelikle veri setimizi görüntüleyelim ve liste elemanlarına bakalım.



Görüntüsü:



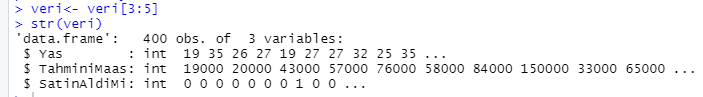
Listemizin Elemanları:



Naive Bayes yöntemi için sayısal sütunları almamız gerekiyor bu yüzden 3:5 sütun aralığını seçiyoruz.



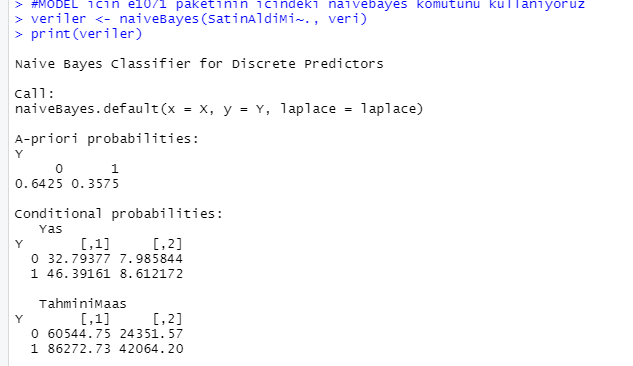
Çıktısı:



Modelimiz için e1071 paketinin içindeki naivebayes komutunu kullanıyoruz.



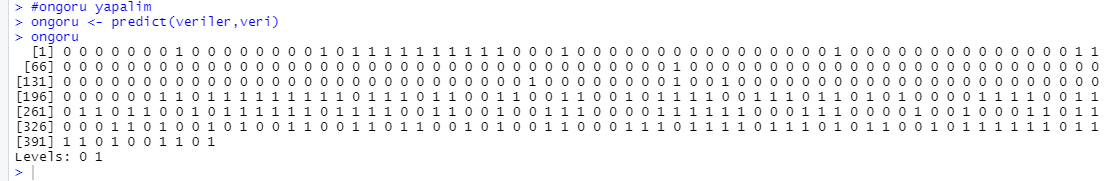
Çıktısı:



Modelimize öngörü yapalım.



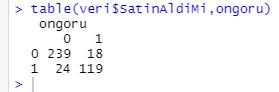
Çıktısı:



Gerçek sınıflandırma değerleri ille öngörü değerlerini karşılaştıralım.



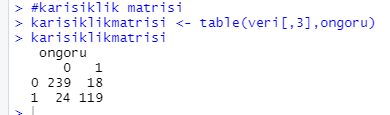
Gördüğünüz gibi gerçekte 239 değer hayır değer varken 18 tanesini yanlış sınıflandırmış. Gerçekte 119 değer evet demişken 24 tanesi yanlış sınıflandırılmış.



Modelimizin karışıklık matrisini görelim.



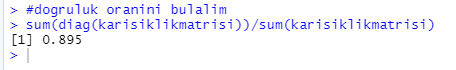
Çıktısı:



Modelimizin doğruluk oranını bulalım.



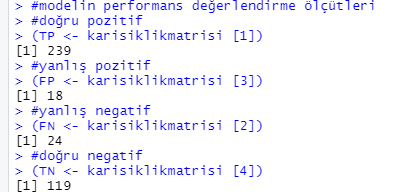
Doğruluk Oranı: 0.895



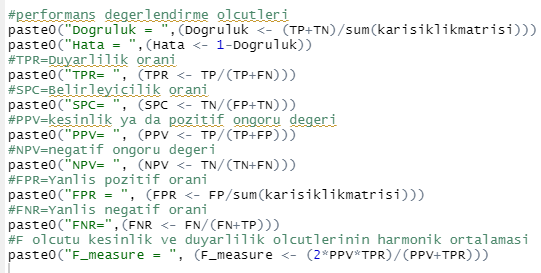
Modelin performans değerlendirme ölçütlerini getirelim.



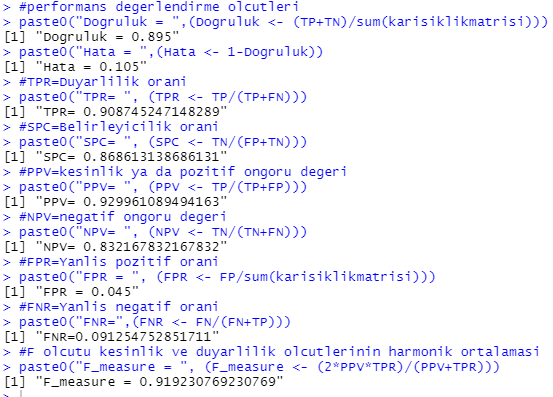
Çıktısı:



Modelin performans değerlendirme ölçütlerini getirelim.



Sonuçları:



**Karar Ağacı İkili Sınıflandırma Matrisi**

|  |  |  |  |
| --- | --- | --- | --- |
| **Gerçek** | | | |
|  |  | 0 | 1 |
| 0 |  | 232 | 25 |
| **Tahmin** 1 |  | 9 | 134 |

**Karar Ağacı İkili Sınıflandırma Algoritmaları Oranları**

|  |  |
| --- | --- |
| **Karar Ağacı Test Seti** | |
| Doğruluk Oranı | %91.5 |
| Hata Oranı | %8 |
| Kesinlik Oranı | %90 |
| Duyarlılık Oranı | %96 |
| Belirleyicilik Oranı | %84 |
| F-Ölçüt | %93 |

**Naive Bayes İkili Sınıflandırma Matrisi**

|  |  |  |  |
| --- | --- | --- | --- |
| **Gerçek** | | | |
|  |  | 0 | 1 |
| 0 |  | 239 | 18 |
| **Tahmin** 1 |  | 24 | 119 |

**Naive Bayes İkili Sınıflandırma Algoritmaları Oranları**

|  |  |
| --- | --- |
| **Karar Ağacı Test Seti** | |
| Doğruluk Oranı | %89 |
| Hata Oranı | %10 |
| Kesinlik Oranı | %92 |
| Duyarlılık Oranı | %90 |
| Belirleyicilik Oranı | %86 |
| F-Ölçüt | %91 |

* 1. **Oluşturulan Modellerin Başarım Ölçütleri**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **TP** | **FP** | **FN** | **TN** | **Doğruluk** | **Hata** | **Kesinlik** | **Duyarlılık** | **Belirleyicilik** | **F-Ölçütü** |
| **Karar Ağacı** | 232 | 25 | 9 | 134 | %91.5 | %8 | %90 | %96 | %84 | %93 |
| **Naive Bayes** | 239 | 18 | 24 | 119 | %89 | %10 | %92 | %90 | %86 | %91 |

Uyguladığımız 2 yöntem kıyaslandığında Naive Bayes’te hem TP(doğru pozitif), hem de FN(yanlış negatif) değeri yüksek çıkmıştır. Karar Ağacında ise FN( yanlış negatif). Hem de TN(doğru negatif) daha yüksek çıkmıştır. Doğruluk oranı, Duyarlılık oranı, F-Ölçütü Karar Ağacı algoritmasında daha yüksek bir orana sahiptir. Hata oranı, Kesinlik oranı, Belirleyicilik oranı değerleri ise Naive Bayes’te daha yüksektir.

1. **Sonuçlar**

Tüketim olgusun içerisinde bulunan satın alma faaliyetlerinde hedef kitlenin karar vermsini etkileyen en önemli unsurlardan biri maaştır. Bu çalışmada tüketicilere ait verileri değerlendirdiğimizde satın alıp almamayı en iyi değerlendiren yöntem Karar Ağacı algoritmasıdır.

Karar Ağacı algoritmasında tüketicinin yaşı ve maaşı ürünü satın alıp almama konusunda etkili unsurlardır. Karar Ağacında hep FP( yanlış pozitif) hem TN ( doğru negatif) değerleri daha yüksek çıkmıştır. Naive Bayes’te TP ve FN değerleri daha yüksek çıkmıştır.

Naive Bayes’te doğruluk oranı %89 iken, Karar Ağacında %91.5’tir. Doğruluk oranına göre değerlendirildiğinde Karar Ağaçlarının daha iyi sonuç verdiğini görmekteyiz.

Duyarlılık oranına bakıldığında Naive Bayes’te %90, Karar Ağacı ise %96 sonucuyla burada da en iyi sonucu vermiştir.

Kesinlik ölçütü dikkate alındığında Karar Ağacı’nın sonucu %90, Naive Bayes’in sonucu %92 çıkmıştır. %92 oran ile en iyi sonucu Naive Bayes vermiştir.

Kesinlik ve duyarlılık ölçütlerinin harmonik ortalaması olan F-ölçütüne bakılarak bu iki ölçüt bir arada değerlendirilmiştir. Bu ölçüte göre karar ağacının %93 oran ile en iyi sonucu verdiğini görmekteyiz. Diğer yandan Naive Bayes’in F-ölçüt değeri %91’dir.

Rakamlar doğrultusunda özetlersek Doğruluk oranı, Duyarlılık oranı, F-Ölçütü Karar Ağacı algoritmasında daha yüksek bir orana sahiptir. Hata oranı, Kesinlik oranı, Belirleyicilik oranı değerleri ise Naive Bayes’te daha yüksektir.

Tüm bu çıkarımlardan sonra Karar Ağacı algoritması diğer algoritmaya göre daha başarılı bulunmuştur.

1. **Kaynaklar**

https://github.com/atacemrah/WoWsignals/commit/5cc959863906feaca2576b6bf078b97a7d96e90c

https://ders.bilecik.edu.tr/pluginfile.php/262302/mod\_resource/content/1/J48kararagaclari.R

<https://ders.bilecik.edu.tr/pluginfile.php/267978/mod_resource/content/1/Naive_Bayes.R>

https://www.ibaness.org/bnejss/2015\_01\_01/06\_torun\_new.pdf

https://medium.com/veri-madencili%C4%9Fi/veri-madencili%C4%9Fi-y%C3%B6ntemleri-cf87b4d40eab

https://avys.omu.edu.tr/storage/app/public/cmumcu/97581/5.%20HAFTA.pptx

https://medium.com/veri-madencili%C4%9Fi/veri-madencili%C4%9Fi-s%C3%BCreci-11243e7966fc

https://dergipark.org.tr/tr/download/article-file/752270

https://hakan.io/karisiklik-matrisi-confusion-matrix/